What affects the tear strength of paperboard?

Consequences of unbalance in a designed experiment

Statistics
Bachelor thesis
Abstract
Supervisor: Jari Appelgren
“What affects the tear strength of paperboard?”

This essay covers a designed experiment on paperboard where the quality under study is tear strength alongside and across.

The objective is to examine what consequences the loss of balance in a designed experiment has on the explanatory power of the proposed empirical model. As did happen, the trial plan didn’t go as planned when the first run caused a disruption of the paperboard in the machine. Decision from the company was to raise the low level of one of the design factors to prevent this from happening again. The consequence of this is an alteration of the design during ongoing experimentation. This in turn affects what analysis approaches are appropriate for the problem.

Three different approaches for analyzing the data are presented, each with different propositions on how to deal with the complication that occurred. The answer to the research question is that the ability of the empirical model to discover significant effects is moderately weakened by the loss of one run (out of eight total). The price payed for retrieving less information from the experiment is that the empirical model, for tear strength across, doesn’t deem the effects significant at the same level as for the candidate model with eight runs. Instead of concluding that the main effect of $B$ and the interaction effect $AB$ is significant at the 2%- and 4%-level, respectively, we must now settle with deeming them significant at the 6%- and 7%-level.

Keywords: Designed experiment, paperboard, statistical analysis, ANOVA, regression, unbalanced design, tear strength
Table of contents

1. Introduction .................................................................................................................. 1
   1.1. Background .............................................................................................................. 1
   1.2. Earlier research on the quality tear strength ......................................................... 1
   1.3. Board Making ......................................................................................................... 1
   1.4. Objective .................................................................................................................. 2
   1.5. Research question ................................................................................................. 2

2. Methodology .................................................................................................................. 3
   2.1. Analysis of variance .............................................................................................. 3
       2.1.1. Fixed effects model .......................................................................................... 3
       2.1.2. Analysis of the fixed effects model ................................................................. 4
   2.1.3. Important concepts in experimental planning ................................................... 5
       2.1.3.1. Basic principles .......................................................................................... 6
       2.1.3.2. Guidelines for designing an experiment ...................................................... 6
   2.1.4. Experimental design for the current problem .................................................... 7
       2.1.4.1. Statement of the problem & objective with experiment ............................... 7
       2.1.4.2. Objective with experiment ......................................................................... 7
       2.1.4.3. Selection of response variable ..................................................................... 7
       2.1.4.4. Choice of factors, levels and ranges ............................................................ 8
       2.1.4.5. Nuisance factors ......................................................................................... 8
       2.1.4.6. Designs ....................................................................................................... 8
       2.1.4.7. Restrictions in experimental planning ......................................................... 11
   2.2. Multivariate Analysis of variance ......................................................................... 11

3. Data ............................................................................................................................... 13
   3.1. Design ..................................................................................................................... 13
   3.2. Data from performed experiment ......................................................................... 14

4. Analysis ......................................................................................................................... 15
   4.1. ANOVA with eight runs ....................................................................................... 15
       4.1.1. Summary ANOVA with eight runs ................................................................. 18
   4.2. ANOVA with seven runs ..................................................................................... 18
       4.2.1. Summary ANOVA with seven runs ............................................................... 21
   4.3. Regression analysis with eight runs ..................................................................... 21
       4.3.1. Summary regression analysis with eight runs ................................................. 22
   4.4. Summary analyses ............................................................................................... 22
   4.5. Multivariate analysis ............................................................................................ 23
   4.6. Candidate model ................................................................................................. 24
4.6.1. Choice of model ........................................................................................................ 24
4.6.2. Model adequacy checking ...................................................................................... 25

5. Discussion ...................................................................................................................... 28
5.1. You can never plan too much .................................................................................... 28
5.2. Unbalanced designs are best avoided ....................................................................... 28
5.3. The conclusions of the experiments alter with the amount of information .......... 28
5.4. Continuing experiments ............................................................................................ 29
List of tables, figures and plots

Table 1 - Example Single-factor Experiment .......................................................... 3
Table 2 - ANOVA table for single factor fixed effects model .................................... 5
Table 3 - Trial plan with randomized runs ............................................................... 13
Table 4 – Coding scheme for the $2^2$ factorial design ........................................... 13
Table 5 - Trial plan .................................................................................................... 14
Table 6 - Data from tear strength experiment (alongside) ......................................... 14
Table 7 - Data from tear strength experiment (across) .............................................. 14
Table 8 - Tear strength data (alongside) ANOVA eight runs .................................... 15
Table 9 - Tear strength data (across) ANOVA eight runs ......................................... 15
Table 10 – Data organization for the two-factor factorial design ............................... 16
Table 11 - ANOVA table for two-factor factorial, fixed effects model ....................... 17
Table 12 - ANOVA analysis eight runs (alongside) ................................................... 17
Table 13 - ANOVA analysis eight runs (across) ........................................................ 18
Table 14 - Tear strength data (alongside) ANOVA seven runs .................................. 18
Table 15 - Tear strength data (across) ANOVA seven runs ..................................... 18
Table 16 - ANOVA analyses seven runs (alongside) ................................................ 20
Table 17 - ANOVA analyses seven runs (across) ..................................................... 20
Table 18 – ANOVA table for regression analysis eight runs (alongside) ..................... 21
Table 19 - Correlation matrix for regression analysis (alongside) ............................. 21
Table 20 – ANOVA table for regression analysis eight runs (across) ......................... 22
Table 21 - Correlation matrix for regression analysis (across) ................................... 22
Table 22 - p-values analyses (alongside) ................................................................. 22
Table 23 - p-values analyses (across) ..................................................................... 23
Table 24 - MANOVA table for tear strength data .................................................... 23
Table 26 - Mean square error for the different empirical models (across) .................... 24
Table 27 - Confidence intervals for effect estimates ANOVA eight runs (across) ....... 25
Table 28 - Confidence intervals for effect estimates ANOVA seven runs (across) ...... 25
Table 29 - Confidence intervals for coefficient estimates regression eight runs (across) .... 25

Figure 1 - $2^2$ factorial design .................................................................................. 9

Plot 1 - MANOVA plot ............................................................................................. 24
Plot 2 - Residual analysis of candidate model ......................................................... 26
Plot 3 - Interaction plot (across) for candidate model ............................................. 26
Plot 4 - Contourplot (across) .................................................................................. 27
1. Introduction

1.1. Background

In the Skoghall Mill in Värmland County of Sweden, Stora Enso produces paperboard to be used in different kind of packages for the food industry. In order to produce a product that satisfies the demand from its customers, the company must ensure that certain qualities are met. The product development division (PDD) is responsible for the development of the paperboard and uses experimental planning as a means to gather insights on what factors influence what qualities.

This essay covers experimental planning and analysis of an important quality of the paperboard, tear strength. During the last year, the quality has degraded without any active measures taken to inflict this. Production has run according to normal procedure. To gain insight into what factors affects the tear strength, a decision was taken from the head of PDD to perform experimental planning and analysis on this quality. It is also the scope for this essay.

1.2. Earlier research on the quality tear strength

Searching through academic journals, I’ve found two papers on the topic tear strength that’s interesting for the scope of this study. They both demonstrate the usefulness of the analysis of variance (ANOVA) method for analyzing the effect different factors might have on the tear strength. Both these works were conducted on textiles, but the application of the analysis method, with its possibilities for implementing powerful visual tools, are independent of the actual material being tested. Especially Asim and Mahmood’s paper has served as an inspiration for this essay in how to present the results of the conducted analysis in an informative and appealing way.¹

1.3. Board Making

Stora Enso mentions in the company folder “Paperboard guide” that²

> The basic principles of paper and paperboard making have not changed for more than two thousand years. Fibres gained from timber are evenly distributed in water. Multiple layers of furnish are applied, one after the other, on a wire. The water is drained from the pulp and the layers are formed into a strong fibre mat. A smooth surface is achieved by coating and calendering.

As for the paperboard properties, there are three important ones; convertibility, printability and protections of content.³ The last one is the one in focus for this experimental planning. Protection of content is about the product not bursting or tearing as well as avoiding box compression. For an overview of the manufacturing process of paperboard at Skoghalls Bruk, I refer to the mentioned Paperboard guide as well as the clip “Skoghalls bruk film Svensk”⁴ that Stora Enso Sverige has uploaded on Youtube.

---

³ Goldszer, Kristian, Stora Enso. Board making and Quality Control, Powerpoint-presentation, slide 2.
⁴ Clip available 2017-09-09.
1.4. Objective

The objective with this essay is to examine what consequences the loss of balance in a designed experiment has on the explanatory power of the empirical model⁵.

1.5. Research question

Does unbalance in the experimental design seriously weaken the ability for the empirical model to discover significant effects?

---

⁵ Models based on actual observations of the system under study. See Montgomery, Douglas C. Design and Analysis of Experiments (2009) p. 2.
2. Methodology

2.1. Analysis of variance

The fundament for the analysis of an experimental design is the analysis of variance (ANOVA). Both the experimental design itself and ANOVA dates back almost hundred years, to the mathematician R.A. Fisher who developed them in agricultural research in the 1920s. In a few decades, by the end of 1950s, Fishers’ statistical methods for experimental research had spread to other fields such as psychology, sociology and engineering. Parolini mentions the profound impact the entry of statistical methods had on experimental practices during the twentieth century: “[Q]ualitative evidence has largely been replaced by quantitative results and the tools of statistical inference have helped foster a new ideal of objectivity in scientific knowledge”.

The following example will shed light upon ANOVA. The table below shows how the data would appear in a single-factor experiment.

Table 1 - Example Single-factor Experiment

<table>
<thead>
<tr>
<th>Treatment level</th>
<th>Observations</th>
<th>Totals</th>
<th>Averages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>y_{11}</td>
<td>...</td>
<td>y_{1}</td>
</tr>
<tr>
<td>2</td>
<td>y_{12}</td>
<td>...</td>
<td>y_{2}</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>a</td>
<td>y_{a1}</td>
<td>...</td>
<td>y_{a}</td>
</tr>
</tbody>
</table>

We have a treatments on different levels of a single factor that we want to compare for possible differences among them. The dots behind the letters or numbers in the cells y_{1}, y_{a}, y_{.} and so forth, is read as summations. The above table and its notations will serve as a reference in the coming account of the empirical model for the data.

2.1.1. Fixed effects model

A normal strategy when faced with analyzing the observations in an experiment is the fixed effects model. The “fixed effects” part comes from that we deliberatively choose which levels the design factors will be set to during the experiment. We know at which range and levels we want to examine if the design factors are significant, thus we set the factors to the levels of our interest. Our goal is to draw conclusions valid only for the levels under study. This is opposed to the random effects model, where the levels of the design factors are randomly chosen. The aim here is to make inferences about the whole population of factor levels.

The definition of the fixed effects model is

\[ y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad \{i = 1, 2, ..., a \} \quad \{j = 1, 2, ..., n \} \tag{2.1} \]

---


7 Table as in table 3.2. in Montgomery, Douglas C. *Design and Analysis of Experiments* (2009), p. 68.

where $\mu$ is the overall mean, common to all the treatments, $\tau_i$ is a parameter unique to the $i$th treatment called the $i$th treatment effect, and $\varepsilon_{ij}$ is a random error component that contain all other sources of variability in the experiment, such as variability transmitted from uncontrolled factors, the experimental units (e.g. variability in the raw material used) and measurement error. Montgomery underlines that the fixed effects model is a suitable model for the experimental design, given that it comes with the features that $\mu$ is a constant and the treatments effects are interpreted as deviations from this constant. A central assumption in the model is that the model errors are assumed to be normally and independently distributed random variables, that is, $N(0, \sigma^2)$. The variance $\sigma^2$ is assumed to be constant for all levels of the factor. This gives that the observations

$$y_{ij} \sim N(\mu + \tau_i, \sigma^2) \tag{2.2}$$

and that the observations are mutually independent. 9

Important to note is that the above equation refers to a one-way, or single-factor analysis of variance, because only one design factor is examined. The equation is expanded with more terms when we consider more design factors, as we shall see further on.

2.1.2. Analysis of the fixed effects model

2.1.2.1. The decomposition of the Total Sum of Squares

The analysis of variance originates from a partitioning of total variability in a dataset into its component parts. The derivation starts off with defining the total corrected sum of squares

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2 \tag{2.3}$$

which is used as the measure for the total variability in the data. Note that the expression matches the numerator in the formula for the sample variance

$$S^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1} \tag{2.4}$$

and hence, the formula for $SS_T$ as a measure of total variability makes sense. Equation 2.3 can, after some manipulations, be written as

$$n \sum_{i=1}^{a} (\bar{y}_i - \bar{y})^2 + \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2 \tag{2.5}$$

As Montgomery mentions, equation 2.5 is the fundamental ANOVA identity. It shows that the total variability in the data can be expressed as, and partitioned into, a sum of squares of the differences between the treatment averages and the grand mean, and a sum of squares of the differences of observations within treatments from the treatment average. The first shall be seen as capturing the differences between the treatment means, and the second as a measure of the random error of the data, since it measures the differences of observations within a treatment from the treatment average. So, in a more mundane manner we could express equation 2.5 as

---

10 The second summation in $SS_T$, as compared to the formula for sample variance, comes from summing over the replicates in the design. Otherwise the formulas would have been identical.
\[ SS_T = SS_{Treatments} + SS_E \] (2.6)

where \( SS_{Treatments} \) captures the sum of squares due to treatments (between treatments), and \( SS_E \) the measure of the sum of squares due to error (within treatments).  \(^{11}\)

### 2.1.2.2. Statistical analysis of the fixed effects model

As applies in any ANOVA analysis, if the between sum of squares is considerably bigger than the error sum of squares, we have stumbled upon a significant treatment effect, that is, we reject the null hypothesis of no treatment effect. Given that the null hypothesis of no treatment effects \( H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0 \) is true, the ratio

\[ F = \frac{SS_{Treatments}/(a-1)}{SS_E/(N-a)} = \frac{MS_{Treatments}}{MS_E} \] (2.7)

is distributed as \( F \) with \((a - 1)\) and \((N - a)\) degrees of freedom. Equation 2.7 is the test statistic for the hypothesis of no difference in treatment means. Now, given that we deal with a \( F \)-distribution, if \( MS_{Treatments} \) significantly differ from \( MS_E \) we shall reject the null hypothesis of no difference in treatment means. For these types of problems, we have a single upper-tail rejection region, and should reject the null hypothesis if

\[ F_0 > F_{\alpha,a-1,N-a} \] (2.8)

where \( F_0 \) is computed from equation 2.7.  \(^{12}\)

To compute the different sum of squares we need to perform the F-test, we look to the following ANOVA table that summarizes everything we need computational-wise to perform the test.  \(^{13}\)

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>( F_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Treatments</td>
<td>( SS_{Treatments} = \sum_{i=1}^{a} \sum_{j=1}^{n} (\bar{y}<em>{ij} - \bar{y}</em>.)^2 )</td>
<td>((a - 1))</td>
<td>( MS_{Treatments} )</td>
<td>( F_0 = \frac{MS_{Treatments}}{MS_E} )</td>
</tr>
<tr>
<td>Error (within treatments)</td>
<td>( SS_{Error} = SS_T - SS_{Treatments} )</td>
<td>((N - a))</td>
<td>( MS_E )</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>( \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_.)^2 )</td>
<td>((N - 1))</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I will make use of corresponding tables in the analysis of the experiment later on in this essay.

### 2.1.3. Important concepts in experimental planning

The following parts under 2.1.3. serve as a short introduction to the topic experimental planning.

---

12 Ibid, p. 73-74.
13 Ibid, p. 75.
2.1.3.1. Basic principles

There are three basic principles in experimental planning:

- **Randomization** - Order of individual runs must be randomly selected to averaging out the effects of extraneous factors.
- **Replication** - Independent repeat run of each factor combination. Allowing for the obtaining of an estimate of the experimental error and more precise estimate of the response parameter (e.g. mean, standard deviation).
- **Blocking** - Used for improving the precision with which comparisons among the factors of interest are made. Blocking should be read as “blocking out” the variability transmitted from nuisance factors, whose potential effect on the response variable we’re not interested in.\(^\text{14}\)

2.1.3.2. Guidelines for designing an experiment

Montgomery introduces a comprehensive set of guidelines for designing an experiment, which I have adopted in this essay. It has seven steps:

1. Recognition of and statement of the problem
2. Selection of the response variable
3. Choice of factors, levels and ranges
4. Choice of experimental design
5. Performing the experiment
6. Statistical analysis of the data
7. Conclusions and recommendations\(^\text{15}\)

Steps 1-4 serves as the design part and steps 5-7 as the analysis part of the experiment. Below, I will shortly discuss steps 1-4. The remaining steps are accounted for by data, analysis and discussion sections in the essay.

Step 1 “Recognition of and statement of the problem” seems somewhat straightforward at first glance, but can be harder than it looks. It isn’t always easy to decide if a problem requires a experimental approach in order to be solved, neither is the problem itself not always easy to define in detail. Montgomery stresses that a clear and generally accepted statement of the problem must be developed, and recommends a team approach to solve this issue. By this means the problem could be illustrated from different perspectives and hence give guidance to how the problem should be stated. In addition to the statement of the problem, the overall objective of the experiment must be decided. Montgomery mentions a handful of different possible objectives: factor screening, optimization, confirmation, discovery and robustness.\(^\text{16}\) Each which affects what type of design that’s appropriate for the problem.

Step 2 of choosing the response variable includes of selecting a response variable that the experimenter is certain about provides useful information of the process under study. Decision of what specific characteristic that should be used to measure the response variable must also be made. The choice often falls on either the mean or the standard deviation, or both. Thought must also be put in securing a reliable measuring system for the response metric, before

---


\(^\text{16}\) Ibid, p. 14f.
conducting the experiment. Calibration of the measurement system and examination of the measurement error of the system could be possible measures to take beforehand.\textsuperscript{17}

Step 3 of choosing factors, levels and ranges includes classifying factors as potential design factors (factors to be included in the experiment) or nuisance factors, and choosing levels and ranges for the chosen design factors. The design factors are chosen based on process knowledge of people working with the system under study. Theoretical and/or prior experience from these experts guides the selection of the factors. Nuisance factors, one can think, shouldn’t be of any interest then, when we have chosen the design factors? Unfortunately, these can have significant effects that must be accounted for. We aren’t interested in them in the context of the present experiment, but the influence that they may have on the result must be considered. The nuisance factors are often classified into three different categories: controllable, uncontrollable and noise factor. The controllable ones are factors where its levels can be set by the experimenter (blocking). The uncontrollable can’t be controlled, but could be measured in the experiment, and its effects can be compensated for by the use of analysis of covariance. Noise factors varies naturally, but can be controlled during the experiment. When having decided which factors are to be the design factors, the experimenter must choose the ranges and levels at which the runs will be made. Once again, the process knowledge from the experts should guide this decision.\textsuperscript{18}

Step 4 of choosing the actual experimental design for the present problem should be relatively straight-forward, given that steps 1-3 has been performed adequately. Things to consider in this step are nonetheless sample size (total number of runs, number of replicates) and if we deal with some kind of randomization restriction (for example blocking).\textsuperscript{19}

2.1.4. Experimental design for the current problem
2.1.4.1. Statement of the problem & objective with experiment

According to the head of the PDD, the tear strength of the produced paperboard at the plant has declined during the last year, with no obvious or major changes in production that could have caused this. The company has a need to find the levels on the influential variables that stabilizes the tear strength on a satisfactory level.

2.1.4.2. Objective with experiment

In discussions I’ve held with the PDD, we have concluded that the objective with the experimental design for the current problem with reduced tear strength is factor screening. The overarching argument is that the division knows too little today of what factors, and possible interactions between factors, that affects the tear strength.

2.1.4.3. Selection of response variable

The response variable for the experiment will be the measured tear strength in milli-Newton (mN) of the produced paperboard. The quality (i.e. the tear strength) is measured in a laboratory environment in the quality assurance division. Standardized tools (Autoline bench) and operational methods will be used, as tear strength is a quality that is inspected as part of the normal quality assurance routine. Each measure of the tear strength is actually a mean of five different measures on the same sample of paperboard. The tear strength is measured both in the direction of the fibers (alongside) and across, so we deal with two response variables in this experiment. The measurements will be taken furthest out on the produced jumbo roll of paper,

\textsuperscript{17} Montgomery, Douglas C. \textit{Design and Analysis of Experiments} (2009), p. 15f.
\textsuperscript{18} Ibid, p. 16f.
\textsuperscript{19} Ibid, p. 18f.
which gives us the best possible guarantee that the changes in levels of the design factors (from one run to another) have materialized on the paperboard. Each run in the experiment is associated with one produced jumbo roll.

2.1.4.4. Choice of factors, levels and ranges

Based on theoretical knowledge of the engineers in the PDD, as well as the operators in manufacturing, decision was made by the head of PDD that the experiment should include two design factors: factor A and factor B.

Factor A today consists of X & Y in a certain proportion. The PDD is interested in what happens if the Y part is increased. This factor will be set to two levels with lower X proportion than the current settings.

Factor B is today set at Z. Given that the objective with the experiment is factor screening, it’s a good idea to widen the gap between the levels of the factor to increasing the chances of being able to tell if it’s a significant and influential factor for the response variable. If the range is set to narrow, we may not be able to tell this as easily. The PDD decided that factor B will be tested at both a lower and higher level than Z.

2.1.4.5. Nuisance factors

A possible nuisance factor in the experiment is the operators in manufacturing. Their job is to monitor and stabilize the manufacturing in daily production. Instructions will be given to the operators that different types of stabilizing or optimizing operations mustn’t be performed during the experiment. These types of operations could otherwise seriously endanger the reliability of the experiment.

Another possible factor that can influence the accuracy in the measured yield, is the time it takes to get full impact from the changed levels in the factors. Here we have to rely on the operators’ knowledge of when changes in the levels of factor A and B fully materializes in the produced paperboard. Their estimate is that it takes 10 minutes to see the effects of changes in factor A and between 20-30 minutes for changes in factor B. Given that the actual measurements will be taken furthest out on the jumbo rolls, which in time corresponds to more than 30 minutes of paperboard production, we are confident that the changes in the levels of the factors have fully materialized on the paperboard.

The wood used for producing the paperboard could also be a possible nuisance factor. The PDD state that wood harvested during different seasons is a source of variation in the manufacturing process. Usage of own pulp or dried pulp (that has to be soaked again) is also a source of variation the affects the yield. Wood from the same season as well as the company’s own pulp will be used in the experiment.

2.1.4.6. Designs

Together with the PDD I’ve discussed what different designs that could be applied to the research problem. Mainly, I introduced full factorial designs and fractional factorial designs, as I saw these as a good introduction to what possible designs that could fit to our research problem. Both designs give, or could give, good estimates of both the main and interaction effects. Definitions of these two effects follows.

2.1.4.6.1. Definitions of main and interaction effects

A main effect is the effect that solely derives from the change in response produced by a change in the level of a factor. An interaction effect occurs when the difference in response between
the levels of one factor isn’t the same for all levels on the other factors, hence the term “interaction”\textsuperscript{20}. An example sheds light upon this subject.

<table>
<thead>
<tr>
<th>Two-factor factorial design (example)</th>
<th>Factor B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low(-)</td>
</tr>
<tr>
<td>Factor A</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>15</td>
</tr>
</tbody>
</table>

The example consists of a $2^2$ full factorial design (more on this later) and can be visualized in the shape of a square:

![Figure 1- $2^2$ factorial design](image)

The computation for the main effects is straight-forward: we evaluate factor $A$ as the difference between the mean for $A^+$ and $A^-$. In this example, the value for the main effect of $A$ is 10.

Main effect of $A$: $A = \bar{y}_{A^+} - \bar{y}_{A^-} = \frac{15+5}{2} - \frac{10+30}{2} = \frac{(15+5-10-30)}{2} = -10$

Main effect of $B$: $B = \bar{y}_{B^+} - \bar{y}_{B^-} = \frac{30+5}{2} - \frac{10+15}{2} = \frac{(30+5-10-15)}{2} = 5$

The computation for the interaction effect is as follows: The effect of factor $A$ is evaluated \textit{at the different levels of factor $B$}, so we look to the table and seeks up $B^+$ and then perform the computation $A^+ - A^-$. The corresponding is then performed for $B^-$. As is evident, this computation evaluates the effect of one factor in the light of another. This example reveals that the interaction effect is of considerable size, when comparing it with the main effects.

Interaction effect AB: $AB = \frac{(5-30)-(15-10)}{2} = \frac{5-30-15+10}{2} = -15$

This example only serves as a brief introduction to the distinction between main and interaction effects. For a more complete account, I refer to Montgomery (2009). In the analysis section of this essay I will give a complete account for the computational methods used.

2.1.4.6.2. Full factorial designs

In a full factorial design, every replicate is complete and every possible combination of the levels of the factors are run, hence the term “full”. In the example above, regarding main and interaction effects, a single-replicate full factorial design was used. We had two factors run at two levels each. It’s common to refer to factorial designs with numbers; in the example that would be $2^2$. Generally, we speak of $N^k$ designs, where the $k$ tells us how many factors that are included in the model, and $N$ how many levels those factors are varied upon.

There are a few advantages with factorial designs that is worth mentioning:

- In a comparison with one-factor-at-a-time (OFAT) experiments$^{21}$, factorial designs are more efficient. We retrieve the same information with fewer runs.
- The design captures possible interaction effects between factors, which OFAT:s does not. It is of major importance to detect one if one exists, we could otherwise state improper conclusions.$^{22}$

2.1.4.6.3. Fractional factorial designs

As the name suggests, in a fractional factorial design not all the runs in a full factorial design are run. The idea of only running a fraction of a full factorial design is based on the assumption that it’s not likely that all higher-order interactions between factors are significant.$^{23}$ $^{24}$ By the means of this assumption we can reduce the number of runs, but still get good estimates of the main effects and low-order interactions of the factors in the design. The fractional factorial sees a major use in factor screening experiments where the goal is to identify which of the factors, among several, that is significant.

The main advantages of a fractional factorial design are:

- The sparsity of effects principle
- If the analysis of a fractional factorial design suggests one of the factors are insignificant, the design can “project” into a larger design in a subset of significant factors.$^{25}$

2.1.4.6.4. Comparison full factorial with fractional factorial designs

In terms of efficiency and economy, the fractional factorial design is superior to the full factorial. The word fractional explicitly declares that not all the possible factor-combinations are run, as compared with the full factorial design which is complete in this sense. The motivation for the fractional factorial design is the scarcity of effects-principle, which comes with the assumption that it isn’t likely that every design factor is significant, and hence, not every single possible factor-combination needs to be run for us to be able to single out these non-significant factors. In comparison with the full factorial design, this reduces the amount of runs we have to perform to reach this conclusion. Hence the greater economy and efficiency.

---

$^{21}$ OFAT designs only varies one factor at a time for each run, making these designs less efficient compared to factorial designs, where all treatment combinations are run. See Mongomery (2009) p. 186f for a more comprehensive account.


$^{23}$ Also called the ”sparsity of effects principle”.

$^{24}$ A higher-order interaction effect is one that the level of several factors makes up for the significant effect (e.g. the high level of A, together with the low level of B, together with the high level of C.

In more technical terms, the fractional factorial design aliases the main effects with higher order interaction effects, i.e. puts a equal sign between these two. It follows that these cannot be separated anymore, and is thus the sacrifice we have to pay for not having to run as many runs as in a full factorial design. The key here is that we get good estimates of the main effects (the scarcity of effects-principle important here!) at reduced cost. As Montgomery emphasizes, experimental planning should be seen as an iterative operation, where findings from one experiment leads to insights on how the next should be conducted, and so on. So, economy in the sense of a factor screening experiment can be read as not wasting runs on factors that later will be judged non-significant. If we can come to the same conclusion using fewer runs, we have more money left to spend for future runs or experiments.

The strength of a full factorial lies in the completeness in the design. All possible factor combinations are run and hence even higher order interactions, if present, will be estimated. Given a proper design, with well-chosen number of levels and ranges for the design factors, we also attain estimates of the effects of one factor over a range of different experimental conditions. These estimates can give good guidance as to in which regions of which factors further experiments should be conducted.

2.1.4.7. Restrictions in experimental planning

An obvious restriction to any experimental planning is economy. Without any restrictions regarding how much the experiment can cost, we could apply designs with a lot of factors and a lot of runs. The design would probably give us a very satisfactory result when it comes to being able to map down the cause-and-effect relationship in the examined system. The power of the F-test of significant differences between the levels of the factors would also be high. But large designs tend to be very costly, and as it happens, the economy have set the limit for the design size in this study. The company could afford sacrificing one day of production for the sake of experimental planning, which corresponds to eight runs. These were the conditions given. Hence, consideration of the power of the tests in the study were not a principle that could guide the size of the design. Important to note is that the mentioned full factorial design make the most use of the observations as possible when it comes to power. In a $2^2$ full factorial we get four observations per level of the different effects we want to estimate ($A$, $B$ and $AB$). This grants the F-test in the ANOVA the best possible requisites to discover truly significant effects.

2.2. Multivariate Analysis of variance

As a complement to ANOVA, and for thoroughness, since we deal with two response variables in this study, I will also evaluate a multivariate approach for the experimental data. The method I will use for analyzing the two response variables together as a group is multivariate ANOVA, or MANOVA. I will not go into detail of the mathematical properties of this method, as this is just a complementary analysis. I refer to Lattin, Carroll & Green (2003) for a comprehensive introduction to MANOVA.

The intuition behind a MANOVA is the same as for the univariate case (ANOVA). Instead of partitioning a scalar sum of squares, we partition a sum of squares matrix. For the two-factor case, we have:

$$S_T = S_A + S_B + S_{AB} + S_E$$  \hspace{1cm} (2.9)

\hspace{2cm} 26 26 Montgomery, Douglas C. Design and Analysis of Experiments (2009), p. 186f.
Instead of using an F-test to evaluate factor and interaction effects, Wilk’s \( \Lambda \) (lambda) statistic are used to compare the sum of squares matrix of the effects, to the sum of squares matrix of the error. The formula for Wilk’s lambda is:

\[
\Lambda = \frac{|S_E|}{|S_T|}
\]  

(2.10)

where \( S_E \) is the residual (or error) sum of squares matrix and \( S_T \) is the total sum of squares matrix. Wilk’s lambda can be modified to test specific hypotheses. Say, for example, that we would like to test for a significant interaction effect. Let \( S_{AB} = S_H \) (H as in “hypothesis”) in the following test statistic:

\[
\Lambda_H = \frac{|S_E|}{|S_H + S_E|}
\]  

(2.11)

When the null hypothesis of no significant effect is true, the ratio will be close to 1.0. When the null hypothesis is false, i.e. we have come across a significant effect, the \( S_H \) part will be large relative to \( S_E \), and the ratio \( \Lambda_H \) will be closer to zero.\(^{27}\)

---

3. Data

3.1. Design

The choice of experimental design by the PDD fell on a $2^2$ full factorial design with two replicates (eight runs). Beforehand, I did a randomization\(^{28}\) of the runs and the resulting trial plan was as follows:

*Table 3 - Trial plan with randomized runs*

<table>
<thead>
<tr>
<th>Trial plan</th>
<th>Factor A</th>
<th>Factor B</th>
<th>Treatment combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run</td>
<td>X %</td>
<td>Y %</td>
<td>Z</td>
</tr>
<tr>
<td>1</td>
<td>60</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>40</td>
<td>200</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>50</td>
<td>200</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>40</td>
<td>200</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>50</td>
<td>200</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

The coding for the treatment combinations in the table above follows the following coding scheme:

*Table 4 – Coding scheme for the $2^2$ factorial design*

<table>
<thead>
<tr>
<th>Treatment Combination</th>
<th>Factorial effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>(1)</td>
<td>-</td>
</tr>
<tr>
<td>a</td>
<td>+</td>
</tr>
<tr>
<td>b</td>
<td>-</td>
</tr>
<tr>
<td>ab</td>
<td>+</td>
</tr>
</tbody>
</table>

The coding scheme rationalizes the notation for the different treatment combinations\(^{29}\). For example, the notation for $A$ low and $B$ low is (1). Should we have computed the different effect estimates by hand, the standard order coding would have helped us a lot. For example, we could have looked to the column “Factorial effect A” and found that this effect is computed as

$$A = \bar{Y}_{A^+} - \bar{Y}_{A^-} = \frac{(ab+a) - (b+1)}{2n} = \frac{(a+ab-(1)-b)}{2n}$$

(3.1)

where the number two in the denominator represents the number of treatment combinations per low and high level of the factors, and $n$ the number of replicates in the design.

The trial plan can also be expressed as in the following table, which is similar to the one presented in the introduction:

---

\(^{28}\) Easily performed in Microsoft Excel with the RAND function.

\(^{29}\) Factor combination and treatment combination are synonyms.
Table 5 - Trial plan

<table>
<thead>
<tr>
<th>Tear strength experiment</th>
<th>Factor B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low (-)</td>
</tr>
<tr>
<td>Factor A</td>
<td></td>
</tr>
<tr>
<td>Low (-)</td>
<td>$y_{111}$</td>
</tr>
<tr>
<td></td>
<td>$y_{112}$</td>
</tr>
<tr>
<td>High (+)</td>
<td>$y_{121}$</td>
</tr>
<tr>
<td></td>
<td>$y_{122}$</td>
</tr>
</tbody>
</table>

3.2. Data from performed experiment

The runs were made during one day of production on one of the production lines in the mill. Unfortunately, due to disruption of the paperboard in the machine (the paperboard broke), the full trial plan couldn’t be performed as planned. The data retrieved from the experiment looks as follows:

Table 6 - Data from tear strength experiment (alongside)

<table>
<thead>
<tr>
<th>Tear strength experiment (alongside)</th>
<th>Factor B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low^{+}(-)</td>
</tr>
<tr>
<td>Factor A</td>
<td></td>
</tr>
<tr>
<td>Low (-)</td>
<td>5479,67</td>
</tr>
<tr>
<td></td>
<td>5493,90</td>
</tr>
<tr>
<td>High (+)</td>
<td>(5439,00)</td>
</tr>
<tr>
<td></td>
<td>5460,20</td>
</tr>
</tbody>
</table>

Table 7 – Data from tear strength experiment (across)

<table>
<thead>
<tr>
<th>Tear strength experiment (across)</th>
<th>Factor B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low^{+}(-)</td>
</tr>
<tr>
<td>Factor A</td>
<td></td>
</tr>
<tr>
<td>Low (-)</td>
<td>6030,8</td>
</tr>
<tr>
<td></td>
<td>6013,77</td>
</tr>
<tr>
<td>High (+)</td>
<td>(6149,00)</td>
</tr>
<tr>
<td></td>
<td>6251,47</td>
</tr>
</tbody>
</table>

Notice that the low level of factor B has been changed from Low to Low^{+}. The mentioned disruption happened during the first run where the treatment combination $a$ was run. Because of this disruption, decision from the engineer in charge was to raise the low level of B to Low^{+}. The consequence of the disruption is thus that one run was run at a different level of $B^-$ than the others (observation in brackets), but we still ended up with eight runs. After the low level was raised for factor B, the trial plan ran without complications.
4. Analysis

The unexpected event during the first run will have implications on what analysis approach should be applied to the problem. One of the strengths with a full factorial design is its complete orthogonality (low level coded as -1 and high level as +1), which is lost when we deal with an unbalanced design, that is, not equal number of observations for every treatment combination (cell in the table above). It also affects the way the sum of squares in an analysis of variance (ANOVA) is computed. I will in the following present three different approaches for analyzing the collected data, each with different assumptions on how to deal with the complication that occurred with the first run.

4.1. ANOVA with eight runs

One approach to the analysis problem is to disregard that the factor-combination \( a \) has two different levels for low for the factor B; Low respectively Low\(^+\). We conduct the analysis as we would have done for a complete design, thinking that the \( a \) runs still consists of estimates for factor A at the high level (\( + \)) and factor B at the low level (\( - \)). The justification for this model is its simplicity.\(^{30}\) We accept that the estimate for \( B^- \) is a combination of two different levels of low. To distinguish this “combined” low level from the general case, I will name it \( B^{mix^-} \), which should be read as “B mixed low”. In this case, we have the following data to analyze:

<table>
<thead>
<tr>
<th>Table 8 - Tear strength data (alongside) ANOVA eight runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>Low (-)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>High (+)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 9 - Tear strength data (across) ANOVA eight runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>Low (-)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>High (+)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Now that we deal with two factors, let’s develop the ANOVA approach first presented in the methodology section 2.1. The organizing of data and the notations we will use are summarized in the table below.

\(^{30}\) Another general justification for an ANOVA model in this type of experimental setting, is that it can be hard to control that the factor reaches the exact factor level as defined in the design, due to a “living” manufacturing environment. For instance, instead of measuring the tear strength at the high level 200 for factor B, the samples contain the levels 198, 202, 204, etc. An ANOVA is more insensitive to this type of departure from the plan than a regression analysis (where the exact level measurement is used), why it can be advised to use an ANOVA in these types of situations.
Expressed mathematically, we have,

\[ y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \quad \bar{y}_{..} = \frac{y_{..}}{abn} \]

\[ y_{i..} = \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \quad \bar{y}_{i..} = \frac{y_{i..}}{bn} \]

\[ y_{.j..} = \sum_{i=1}^{a} \sum_{k=1}^{n} y_{ijk} \quad \bar{y}_{.j..} = \frac{y_{.j..}}{bn} \]

And the grand total and grand average are

\[ y_{...} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \quad \bar{y}_{...} = \frac{y_{...}}{abn} \quad (4.1) \]

If we partition the total corrected sum of squares for a two-factor factorial, we have the following fundamental ANOVA identity:

\[ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \bar{y}_{..})^2 = bn \sum_{i=1}^{a} (\bar{y}_{i..} - \bar{y}_{..})^2 + an \sum_{j=1}^{b} (\bar{y}_{.j..} - \bar{y}_{..})^2 + n \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \bar{y}_{ij..})^2 \quad (4.2) \]

What we see on the right-hand side is the sum of squares due to factor \( A \), factor \( B \), the interaction and the error, respectively. The next step to create computing formulas is straight-forward:

\[ SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn} \quad (4.3) \]

\[ SS_A = \frac{1}{bn} \sum_{i=1}^{a} y_{i..}^2 - \frac{y_{..}^2}{abn} \quad (4.4) \]

\[ SS_B = \frac{1}{an} \sum_{j=1}^{b} y_{.j..}^2 - \frac{y_{..}^2}{abn} \quad (4.5) \]
It is handy to compute the sum of squares for $AB$ in two steps. First, we compute the sum of squares between the $ab$ cell totals, which we name “subtotals”:

$$SS_{Subtotals} = \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij}^2 - \frac{\bar{y}_{..}^2}{abn}$$ (4.6)

$SS_{Subtotals}$ contains $SS_A$ and $SS_B$, so step two to get the estimate for $SS_{AB}$ is to subtract these estimates from $SS_{Subtotals}$:

$$SS_{AB} = SS_{Subtotals} - SS_A - SS_B$$ (4.7)

Finally, we compute the sum of squares for error by subtracting $SS_{Subtotals}$ from $SS_T$:

$$SS_E = SS_T - SS_{Subtotals}$$ (4.8)

The ANOVA approach for a two-factor, fixed effects model is summarized below.

Table 11 - ANOVA table for two-factor factorial, fixed effects model

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>$F_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Treatments</td>
<td>$SS_A$</td>
<td>$a - 1$</td>
<td>$MS_A = SS_A/(a - 1)$</td>
<td>$F_0 = \frac{MS_A}{MS_E}$</td>
</tr>
<tr>
<td>B Treatments</td>
<td>$SS_B$</td>
<td>$b - 1$</td>
<td>$MS_B = SS_B/(b - 1)$</td>
<td>$F_0 = \frac{MS_B}{MS_E}$</td>
</tr>
<tr>
<td>Interaction</td>
<td>$SS_{AB}$</td>
<td>$(a - 1)(b - 1)$</td>
<td>$MS_{AB} = SS_{AB}/(a - 1)(b - 1)$</td>
<td>$F_0 = \frac{MS_{AB}}{MS_E}$</td>
</tr>
<tr>
<td>Error (within treatments)</td>
<td>$SS_E$</td>
<td>$ab(n - 1)$</td>
<td>$MS_E = SS_E/ab(n - 1)$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$SS_T$</td>
<td>$abn - 1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The actual analyses in this essay will be conducted in the programming language R. I will consistently print out only the output of these analyses. The full R code is found in appendix 1.

Table 12 - ANOVA analysis eight runs (alongside)

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>144.76</td>
<td>144.76</td>
<td>0.02</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>6703.08</td>
<td>6703.08</td>
<td>0.71</td>
</tr>
<tr>
<td>A:B</td>
<td>1</td>
<td>1644.80</td>
<td>1644.80</td>
<td>0.18</td>
</tr>
<tr>
<td>Residuals</td>
<td>4</td>
<td>37559.64</td>
<td>9389.91</td>
<td></td>
</tr>
</tbody>
</table>

---

The ANOVA table shows that none of the effect estimates are significant at the usual levels of significance (1-10%). No matter the settings on factor $A$ or $B$ the response is influenced in any significant way.

Table 13 - ANOVA analysis eight runs (across)

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>7675.61</td>
<td>7675.61</td>
<td>1.75</td>
<td>0.2568</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>62881.22</td>
<td>62881.22</td>
<td>14.31</td>
<td>0.0194</td>
</tr>
<tr>
<td>A:B</td>
<td>1</td>
<td>26912.00</td>
<td>26912.00</td>
<td>6.13</td>
<td>0.0686</td>
</tr>
<tr>
<td>Residuals</td>
<td>4</td>
<td>17574.00</td>
<td>4393.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The ANOVA table reveals that the main effect of factor $B$ is significant at the 5%-level and that the $AB$ interaction effect is significant at the 10%-level. The main effect of $A$ is not significant at either of the mentioned significance levels.

4.1.1. Summary ANOVA with eight runs

Overall, we conclude that main effect of factor $A$ is a long way from being significant for either of the two response variables.\textsuperscript{32} Main effect of factor $B$ is highly significant for tear strength across, but not for tear strength alongside, though it’s still the effect with the lowest p-value for this response too. The interaction effect $AB$ is significant for tear strength across at the 10%-level, but not for tear strength alongside.

4.2. ANOVA with seven runs

A different approach to handle the situation with having two different $B$ levels, is to discard the diverging run to get a coherent design with respect to the levels. We’re also avoiding to group two different levels of low together. The loss, or cost, is that we lose one observation. In this case, we would have the following data to analyze:

Table 14 - Tear strength data (alongside) ANOVA seven runs

<table>
<thead>
<tr>
<th>Tear strength data (alongside)</th>
<th>Factor B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor A</td>
<td></td>
</tr>
<tr>
<td>50/50</td>
<td>5479.67</td>
</tr>
<tr>
<td></td>
<td>5493.90</td>
</tr>
<tr>
<td>60/40</td>
<td>5460.20</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 15 - Tear strength data (across) ANOVA seven runs

<table>
<thead>
<tr>
<th>Tear strength experiment (across)</th>
<th>Factor B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
</tr>
<tr>
<td>Factor A</td>
<td></td>
</tr>
<tr>
<td>50/50</td>
<td>6030.8</td>
</tr>
<tr>
<td></td>
<td>6013.77</td>
</tr>
<tr>
<td>60/40</td>
<td>6251.47</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{32} At the usual levels of significance at 1-10%.
In the case of an unbalanced design, proper modifications must be made in how we compute the different sum of squares. The effects estimates are now computed as follows:

\[ SS_A = \sum_{i=1}^{a} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N} \]  

(4.9)

\[ SS_B = \sum_{j=1}^{b} \frac{y_{.j}^2}{n_j} - \frac{y_{..}^2}{N} \]  

(4.10)

\[ SS_{Subtotals} = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{y_{ij}^2}{n_{ij}} - \frac{y_{..}^2}{N} \]  

(4.11)

where \( n_i \), \( n_j \) and \( n_{ij} \) means the number of observations per row, column and specific cell, respectively.

\[ SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{..}^2}{N} \]  

(4.12)

\[ SS_E = SS_T - SS_{Subtotals} \]  

(4.13)

\[ SS_{AB} = SS_{Subtotals} - SS_A - SS_B \]  

(4.14)

In comparison with the computational formulas for the balanced two-factor factorial, the above formulas have been modified to consider that the cells don’t have equal number of observations. One must note that the computational formulas above produces an ANOVA where the main effects of \( A \) and \( B \) will be estimated first. The interaction effect will reduce the residual SS only for the variance not already explained by the main effects of \( A \) and \( B \). This approach thus gives preference to the estimation of the main effects, and should be understood as a type of type I SS (see below).

There are actually a few different types of sum of squares that can be used for the computation of the effect estimates in an ANOVA. This is important to recognize when we deal with an unbalanced design, because they produce different results. I will here discuss two, type I and type III, which sheds light upon the subject of why it’s preferable to work with a balanced design.

Type I SS gives weight to the variables according to the order in which they’re entered in the model, also known as the “sequential” decomposition. So, if we start alphabetically factor \( A \) will have full space to account for as much of the variance in the model as it can. Factor \( B \) on the other hand will only be able to account for the variance that hasn’t already been accounted for by factor \( A \), and so forth. This situation does not occur in a balanced design, given its orthogonality. When we have uncorrelated factors, it doesn’t matter in which order we enter them in the model, they don’t “cannibalize” on each other. It’s worth noting that R uses type I as the default SS.

Type III offers another approach to compute the SS. Instead of giving weight to the order in which the factors are inserted, the type III handles every factor as should it get inserted last, when all other factors that don’t contain the factor already are accounted for in the model. One obvious advantage it has for the problem we face here, compared to type I, is that it will produce the same result regardless of the order the factors are inserted in the model. Hence, it will not
“overestimate” a single factor to the prize of underestimation of another one. The type III SS is considered useful for an unbalanced design with no missing cells because it’s independent of sample size: the effect estimates are not a function of the frequency of observations in any group, that is, the means are weighted equally regardless of how many observations there is in the groups. Finally, an implication of the “last in” approach of the type III is that the SS for the effects and the residuals won’t sum to the total SS for the model, which it always will in the type I SS.33

In the tables below the output from the three type I analyses (A, B and AB inserted first in the analysis, respectively) and the type III analysis for the design are displayed for evaluation.

Table 16 - ANOVA analyses seven runs (alongside)

<table>
<thead>
<tr>
<th>Source</th>
<th>Type I A (A, B, AB)</th>
<th>Type I B (B, A, AB)</th>
<th>Type I AB (AB, A, B)</th>
<th>Type I AB (AB, B, A)</th>
<th>Type III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sum Sq</td>
<td>Pr(&gt;F)</td>
<td>Sum Sq</td>
<td>Pr(&gt;F)</td>
<td>Sum Sq</td>
</tr>
<tr>
<td>A</td>
<td>153</td>
<td>0.92</td>
<td>4</td>
<td>0.99</td>
<td>89</td>
</tr>
<tr>
<td>B</td>
<td>3827</td>
<td>0.62</td>
<td>3976</td>
<td>0.61</td>
<td>4426</td>
</tr>
<tr>
<td>AB</td>
<td>874</td>
<td>0.81</td>
<td>874</td>
<td>0.81</td>
<td>340</td>
</tr>
<tr>
<td>Residuals</td>
<td>37335</td>
<td>0.01</td>
<td>37335</td>
<td>0.01</td>
<td>37335</td>
</tr>
</tbody>
</table>

As expected, we retrieve quite the different SS for the effects from the different type I analyses. Depending on in which order we choose to insert the factors, the SS changes. This very clearly visualizes the implications of using type I SS for an unbalanced design. We can also conclude that the type III analysis contains the same SS for the factors, as in any of the type I analyses when they were inserted last. This is in accordance to what’s been stated above about the type III approach. In comparison with the type I approach, every SS for the factors gets penalized equally, instead of this penalty getting weighted by the (reverse) order in which the factors were inserted. Given this reasoning, I will use the type III SS in the ANOVA with seven runs.

The output from the performed ANOVA above shows that none of the effects are significant for tear strength alongside. This is completely in line with the conclusions from the ANOVA with eight runs.

Table 17 - ANOVA analyses seven runs (across)

<table>
<thead>
<tr>
<th>Source</th>
<th>Type I A (A, B, AB)</th>
<th>Type I B (B, A, AB)</th>
<th>Type I AB (AB, A, B)</th>
<th>Type I AB (AB, B, A)</th>
<th>Type III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sum Sq</td>
<td>Pr(&gt;F)</td>
<td>Sum Sq</td>
<td>Pr(&gt;F)</td>
<td>Sum Sq</td>
</tr>
<tr>
<td>A</td>
<td>13654</td>
<td>0.17</td>
<td>5850</td>
<td>0.32</td>
<td>23170</td>
</tr>
<tr>
<td>B</td>
<td>54013</td>
<td>0.04</td>
<td>61817</td>
<td>0.03</td>
<td>36819</td>
</tr>
<tr>
<td>AB</td>
<td>32089</td>
<td>0.07</td>
<td>32089</td>
<td>0.07</td>
<td>39766</td>
</tr>
<tr>
<td>Residuals</td>
<td>12324</td>
<td>0.01</td>
<td>12324</td>
<td>0.01</td>
<td>12324</td>
</tr>
</tbody>
</table>

Reid, N., University of Toronto: http://www.utstat.utoronto.ca/reid/sta442f/2009/typeSS.pdf (2017-08-18);
The conclusions drawn from this analysis is similar to the ANOVA with eight runs; the main effect of $B$ is significant at the 10%-level (5%-level for ANOVA with eight runs) and the $AB$ interaction effect at the 10%-level for tear strength across. The main effect of $A$ is not significant at any of these significance levels.

4.2.1. Summary ANOVA with seven runs

Overall, we recognize that the main effect of factor $A$ is a long way from being significant for either of the two response variables. Main effect of factor $B$ and the interaction effect $AB$ is significant at the 10%-level for tear strength across, but not for tear strength alongside.

4.3. Regression analysis with eight runs

A common coding of the levels of the factors in a two-level factorial design, when preparing the data for a regression analysis, is to code the low level as -1 and the high as +1. This grants full orthogonality among the contrasts, by which the assumption of uncorrelated regression coefficients holds. This in turn makes the assumption of mutually independent observations hold. 34

The implication of having two different levels of low for factor $B$ is the need to change the coding accordingly. Since one $B^-$ run had the level Low, we code this one as -2. It’s proportional to the coding of the other levels: $B^-$ with Low + gets -1 and $B^+$ with High + gets +1. The result of this coding is that we lose full orthogonality. The implication of this is that the variables get correlated and we could face a multicollinearity problem. Hence, performing this analysis, we must check for correlation among our variables (the factors).

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>144.76</td>
<td>144.76</td>
<td>0.02</td>
<td>0.9070</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>7579.96</td>
<td>7579.96</td>
<td>0.81</td>
<td>0.4189</td>
</tr>
<tr>
<td>A:B</td>
<td>1</td>
<td>909.16</td>
<td>909.16</td>
<td>0.10</td>
<td>0.7708</td>
</tr>
<tr>
<td>Residuals</td>
<td>4</td>
<td>37418.39</td>
<td>9354.60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 19 - Correlation matrix for regression analysis (alongside)

<table>
<thead>
<tr>
<th></th>
<th>resp</th>
<th>A</th>
<th>B</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>resp</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>-0.06</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.41</td>
<td>-0.11</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>0.24</td>
<td>-0.11</td>
<td>0.26</td>
<td>1.00</td>
</tr>
</tbody>
</table>

I’ve transformed the linear model to an ANOVA model for full comparability with the ANOVA analyses. 35 The ANOVA output for the linear model above shows that none of the effects are significant. The same conclusion was drawn in both the previous analyses. The correlation matrix confirms that we don’t deal with severely correlated variables. It also reveals that the main effect $B$ correlates strongest with the response variable (resp), which is in line with the findings in the ANOVA table.

35 Translating the linear model to an ANOVA in R using anova(“my_regr_model”), will generate an ANOVA output based on type I SS.
The summary of the linear model above shows that both main effect of factor $B$ and interaction effect $AB$ is significant at the 5%-level. The main effect of $A$ is not significant at either of the mentioned significance levels. This is similar to the conclusions drawn from the ANOVA with eight runs, with the difference that in the latter the interaction effect was significant at the 10%-level. The correlation matrix confirms that the variables are not correlated to any significant extent.

4.3.1. Summary regression analysis with eight runs

Overall, the analysis show that main effect $A$ is far from being significant for either of the response variables. Main effect $B$ is highly significant for tear strength across, but not for tear strength alongside, though it’s still the effect with the lowest p-value for this response too. The interaction effect $AB$ is significant for tear strength across at the 5%-level, but not for tear strength alongside.

4.4. Summary analyses

Inspecting the summary tables below, one can quickly conclude that, regardless of which analysis method chosen, none of the design factors are significant for the response variable tear strength alongside. Turning to the response variable tear strength across, we conclude that main effect $B$ and interaction effect $AB$ are significant in all the analyses, if we accept a significance level of ten percent. Should we set the level at five percent, only the regression model with eight runs would judge both the effects significant. The ANOVA with seven runs would treat both as non-significant.

**Table 20 – ANOVA table for regression analysis eight runs (across)**

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>7675.61</td>
<td>7675.61</td>
<td>2.03</td>
<td>0.2275</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>58225.33</td>
<td>58225.33</td>
<td>15.39</td>
<td>0.0172</td>
</tr>
<tr>
<td>A:B</td>
<td>1</td>
<td>34005.32</td>
<td>34005.32</td>
<td>8.99</td>
<td>0.0400</td>
</tr>
<tr>
<td>Residuals</td>
<td>4</td>
<td>15136.57</td>
<td>3784.14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 21 - Correlation matrix for regression analysis (across)**

<table>
<thead>
<tr>
<th></th>
<th>resp</th>
<th>A</th>
<th>B</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>resp</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.26</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.68</td>
<td>-0.11</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>-0.37</td>
<td>-0.11</td>
<td>0.26</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Table 22 - p-values analyses (alongside)**

<table>
<thead>
<tr>
<th>Effect</th>
<th>ANOVA 8 runs</th>
<th>ANOVA 7 runs</th>
<th>Regression 8 runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.907</td>
<td>0.973</td>
<td>0.907</td>
</tr>
<tr>
<td>B</td>
<td>0.446</td>
<td>0.593</td>
<td>0.419</td>
</tr>
<tr>
<td>A:B</td>
<td>0.697</td>
<td>0.808</td>
<td>0.771</td>
</tr>
</tbody>
</table>
It’s worth noting that all the analysis methods lead us to draw the somewhat same conclusion; that main effect $B$ and interaction effect $AB$ are interesting for the yield in tear strength across. That said, we recognize that the p-values are further from being significant for all effects for both response variables in the ANOVA model with seven runs, compared to the other analyses.

4.5. Multivariate analysis

For reasons of thoroughness I will also evaluate a multivariate analysis approach to the experimental data. In the light of the ANOVA:s performed, one could expect that the results shown in the univariate cases are reflected in the MANOVA. Generally, an important motivation for running a MANOVA instead of several ANOVA:s are that the univariate analyses cannot take into account possible covariation among the response variables. This can create a situation where a significant treatment effect across the response variables exists, but are not detected by the single ANOVA:s.\textsuperscript{36} We have already found significant effects in the univariate analyses for the response variable tear strength across, why it is expected that this result is reflected in a MANOVA on our experimental data.

In the table below the results from the MANOVA are shown. We conclude that main effect of factor $B$ is significant at the 5%-level for the response variable consisting of tear strength alongside and across together.

<table>
<thead>
<tr>
<th>Effect</th>
<th>ANOVA 8 runs</th>
<th>ANOVA 7 runs</th>
<th>Regression 8 runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.257</td>
<td>0.182</td>
<td>0.228</td>
</tr>
<tr>
<td>B</td>
<td>0.019</td>
<td>0.058</td>
<td>0.017</td>
</tr>
<tr>
<td>A:B</td>
<td>0.069</td>
<td>0.068</td>
<td>0.040</td>
</tr>
</tbody>
</table>

The plot below depicts how the observations are scattered in the two-dimensional space of the response. The high level of $B$ (blue dots) is generally related to a higher yield in the response.

Evaluating the MANOVA in the light of the univariate ANOVA:s, the results are as expected. Main effect of factor A is not significant on the whole, which is perfectly in line with the ANOVA:s. Main effect of B were not significant for tear strength alongside, it was the effect with the lowest p-value of all the effects though, and highly significant for tear strength across. The MANOVA deem it as significant at the 5%-level on the whole, which seems reasonable given the ANOVA findings. The interaction effect AB were a long way from being significant for tear strength alongside, but significant for tear strength across on the 5%-level. The MANOVA deems it significant at the 12%-level, which can be viewed as a compromise between the results in the univariate analyses.

4.6. Candidate model

4.6.1. Choice of model

When coming to the decision of choosing our candidate model for this problem, we should first acknowledge that none of the applied analysis methods show significant results for the response variable tear strength alongside. Turning to the other response variable, tear strength across, we see that all the analysis methods exhibits significant results. Given this, we should choose the one that captures the reality in the manufacturing conditions the best. This model contains the most possible information about the experimental environment that were present when the experiment was run. Given this reasoning I will choose the regression model as my candidate model. This model also has the lowest mean square error (MSE) among the contestants, which gives us a measure of which model that best explains the variation in the data.

<table>
<thead>
<tr>
<th>ANOVA 8 runs</th>
<th>ANOVA 7 runs</th>
<th>Regression 8 runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>66.28</td>
<td>64.09</td>
</tr>
</tbody>
</table>

We can also look to the different effect estimates and their corresponding 95% confidence intervals to support us in our decision of our candidate model:
It’s clear from the above tables that the most precise estimates of the factor effects are made in the regression with eight runs, since it has the tightest interval for the estimates. The widest intervals are found in the ANOVA with seven runs.

4.6.2. Model adequacy checking

An important part of model fitting is to check if the residuals verifies the model assumptions of independently distributed observations. As seen in the graph below, the normality assumption of the residuals holds, as well as they exhibit no sign of heteroscedasticity. We also conclude that all the observations have about the same leverage with no obvious outliers. Thus, we conclude the observations in our experiment are independently distributed.
Let’s turn to what this model actually tells us about the relationship between the factors and the response variable. An interaction plot is a convenient tool to use to display how the factors interact with the response variable. The interaction plot for the candidate model suggests a strong interaction between the factor $A$ and factor $B$, which is most clearly depicted in the difference in the mean response for different levels of factor $B$ when factor $A$ is held at low. At the same time, the plot suggests that the mean response barely is influenced by a change of level in factor $B$, when factor $A$ is held at high. The highest yield is reached when factor $A$ is set to low and factor $B$ is set to high.
Another good tool for visualizing the factors influence on the response variable is a contourplot. It shows in an even better way the expected yield in Y (the tear strength) for the whole range of the factors in the experiment. The plot below displays for instance that the highest yield is reached when factor A is set to the high level and factor B is set to the low level, here the response is around the 6300 mark. We can also conclude that a tear strength of 6250 can be reached for the whole range of factor B, if factor A is set to the level 0,5. Moreover, to reach a tear strength of approximately 6200, the contourplot suggests that this level of yield can be reached by setting factor A to the level of approximately 0,25 and factor B to the high level.

*Plot 4 - Contourplot (across)*
5. Discussion

5.1. You can never plan too much

The experiment didn’t went exactly as planned, even though the pre-planning included addressing the sources of unwanted variation that we could identify prior to conducting the experiment. One obvious improvement would have been to for me to attend the experiment, which I couldn’t do out of timely reasons. Given that I was given the task to develop the experimental design, me together with the engineer in charge and the operators could have had a discussion regarding how we should run the remaining part of the experiment when the first run failed. The alteration of the low level of factor B from Low to Low+ was a required step to take, given that Low caused disruption of the machine. The trial plan could then have been restarted with the first run, now at Low+, and then continue on according to the new plan. The main reason for this is that we would have gotten at completely orthogonal and computationally smooth model if all eight runs would have run as planned. Now when this couldn’t be fulfilled, we faced a problem which called for slightly different analysis methods to be employed.

The fact is, though, that what was later chosen as the candidate model didn’t require much extra-work compared with the preparations of a full orthogonal model. A slightly different coding and the need to check for potentially correlated variables were the steps added. The regression model approach proved to be very useful in this account, which Montgomery also acknowledge.37 One has to bear in mind though, that we still got our eight runs, otherwise we would have had to deal with a design with seven runs.

5.2. Unbalanced designs are best avoided

As it turned out, we still had eight runs for our disposal at the end of the experiment, even though one didn’t turn out exactly as planned. I’ve shown the benefits of being able to use this observation (in what became the candidate model). But what would have been the implications if we only had obtained seven runs? In the analysis section, I’ve given account for the choices we face in this type of situation. Turning to a model based on type I SS, we stand a choice of subjectively choose in which order to insert the factors into the model, knowing that it will have implications on the estimations of the SS of the effects. Say we face a situation where several of the factors dangling around the 5%-level of significance, what would our conclusion be? We could try fitting different models where we for each one let a different factor be inserted first, to get estimations of the SS for the effects at their most favorable position. We then have to somehow make a decision of how to interpret and evaluate the findings from our experiment. Or we could do as I did in this essay, and choose to run with the type III SS, knowing that we by this approach get more conservative estimations. In both cases, I believe it’s not far-fetched to think that we would come to the conclusion that it would be good to complement the performed experiment with more runs. This would give us more precise estimates of the effects. In the ANOVA with seven runs one of the treatment combinations only had a single replicate. The implication is that we lost the estimate of experimental error in this treatment combination, making it harder to detect significant effects in our model.38

5.3. The conclusions of the experiments alter with the amount of information

The answer to the research question is that the conclusions drawn from the experiment are moderately weakened by the loss of a run. The price payed for retrieving less information from the experiment is that the empirical model doesn’t deem the effects significant at the same level

38 Ibid, p. 12.
as for the candidate model with eight runs. The result stands true for both response variables. Instead of concluding that the main effect of $B$ and the interaction effect $AB$ is significant at the 2%- and 4%-level, respectively, for tear strength across, we must now settle with deeming them significant at the 6%- and 7%-level. If the experimenter isn’t willing to risk more than a five percent chance of randomness playing a part in us retrieving evidence of significant effects, we would come to two different conclusions: in the model with eight runs we have two significant effects, in the model with seven runs we have none.

Of course, the estimates would still be very close to the limit where the experimenter would regard the effects as significant. When the results imply interesting relationships between the design factors and the response, it’s not far-fetched to think that the experimenter would want to run a few more replicates to be able to obtain more precise estimates of the effects. This would of course mean taking the machines in possession once more and performing complementary runs. This in turn would leave us with a costlier experiment than would we have retrieved the data we needed for the planned (balanced) design in the first place.

So, the implication of losing information in the form of a discarded run leaves us with more imprecise estimates of the effects, which in turn calls for actions to retrieve more information. Alternatively, we can decide to accept the results of the analysis, knowing that they’re inflicted with more insecure estimates.

5.4. Continuing experiments

The possible continuation of the experiment, in light of the conclusions drawn, must be decided based on the interest of the company. Are they content enough with the results shown this far, and want to investigate even further improvements of the response variable? Then, we should consider moving the “design square” into a new area and new levels for the design factors, where we anticipate an even higher yield. Or are they satisfied with the answers the design gave and want to move on to trying other possible design factors, with or without the ones already used? Then we should draw up a new experimental design. The answer to the above questions, and others, should guide how the next experiment is set up.
Sources

Literature


Academic articles


Web sources


Lane, M. David, Rice University, “*Unequal sample sizes*” (2017-08-22) URL-address: [http://onlinestatbook.com/2/analysis_of_variance/unequal.html](http://onlinestatbook.com/2/analysis_of_variance/unequal.html)


Other

Appendix 1 – R-code

data_al<-c(5479.67, 5439.00, 5559.57, 5406.87, 5493.90, 5460.20, 5472.43, 5665.47)
data_ac<-c(6030.8, 6149.00, 6354.67, 6329.10, 6013.77, 6251.47, 6276.53, 6194.00)

## Table 12 & 13 - ANOVA tables for analysis with eight runs

A<-gl(2,1,8)
B<-gl(2,2,8)
AB<-A*B

fm<-aov(data_al~A*B)
fm2<-aov(data_ac~A*B)

summary(fm); summary(fm2)

## Table 16 - ANOVA tables for analyses with seven runs (alongside)

library(car)
library(Hmisc)
data_across<-c(6030.8, 6354.67, 6329.10, 6013.77, 6251.47, 6276.53, 6194.00)
A<-c(-1,-1,1,1,-1,-1,1)
B<-c(-1,1,1,-1,-1,1,1)
AB<-A*B

A_1st<-aov(data_across ~ A + B + AB)
B_1st<-aov(data_across ~ B + A + AB)
AB_1st<-aov(data_across ~ AB + A + B)
AB_1st_2<-aov(data_across ~ AB + B + A)

typeIII<-aov(data_across ~ A + B + AB)
aov_typeIII<-Anova(typeIII, type="III")

fr_A<-data.frame(c("A", "B", "AB", "Residuals"), summary(A_1st)[1][["Sum Sq"]],summary(A_1st)[1][["Pr(>F)"]])
colnames(fr_A)<-c("Source", "Sum Sq", "Pr(>F)")
fr_B<-data.frame(c("B", "A", "AB", "residuals"), summary(B_1st)[1][["Sum Sq"]],summary(B_1st)[1][["Pr(>F)"]])
colnames(fr_B)<-c("Source", "Sum Sq", "Pr(>F)")
fr_AB<-data.frame(c("AB", "A", "B", "Residuals"), summary(AB_1st)[1][["Sum Sq"]],summary(AB_1st)[1][["Pr(>F)"]])
colnames(fr_AB)<-c("Source", "Sum Sq", "Pr(>F)")
fr_AB2<-data.frame(c("AB", "B", "A", "Residuals"), summary(AB_1st_2)[1][["Sum Sq"]],summary(AB_1st_2)[1][["Pr(>F)"]])
colnames(fr_AB2)<-c("Source", "Sum Sq", "Pr(>F)")
fr_III<-data.frame(c("A", "B", "AB", "Residuals"), aov_typeIII[[-1,1], aov_typeIII[[-1,4]])
colnames(fr_III)<-c("Source", "Sum Sq", "Pr(>F)")
full_fr<-cbind(fr_A, fr_B, fr_AB, fr_AB2, fr_III)
full_fr

## Table 17 - ANOVA tables for analyses with seven runs (across)

library(car)
library(Hmisc)
data_along<-c(5479.67, 5559.57, 5406.87, 5493.90, 5460.20, 5472.43, 5665.47)
A<-c(-1,1,-1,1,-1,1)
B<-c(-1,1,1,-1,1,1)
AB<-A*B
A_1st<-aov(data_along ~ A + B + AB)
B_1st<-aov(data_along ~ B + A + AB)
AB_1st<-aov(data_along ~ AB + A + B)
AB_1st_2<-aov(data_along ~ AB + B + A)
typeIII<-aov(data_along ~ A + B + AB)
aov_typeIII<-Anova(typeIII, type=c("III"))
fr_A<-data.frame(c("A", "B", "AB", "Residuals"), summary(A_1st)[1][["Sum Sq"]], summary(A_1st)[1][["Pr(>F)"]])
colnames(fr_A)<-c("Source", "Sum Sq", "Pr(>F)")
fr_B<-data.frame(c("B", "A", "AB", "Residuals"), summary(B_1st)[1][["Sum Sq"]], summary(B_1st)[1][["Pr(>F)"]])
colnames(fr_B)<-c("Source", "Sum Sq", "Pr(>F)")
fr_AB<-data.frame(c("AB", "A", "B", "Residuals"), summary(AB_1st)[1][["Sum Sq"]], summary(AB_1st)[1][["Pr(>F)"]])
colnames(fr_AB)<-c("Source", "Sum Sq", "Pr(>F)")
fr_AB2<-data.frame(c("AB", "B", "A", "Residuals"), summary(AB_1st_2)[1][["Sum Sq"]], summary(AB_1st_2)[1][["Pr(>F)"]])
colnames(fr_AB2)<-c("Source", "Sum Sq", "Pr(>F)")
fr_III<-data.frame(c("A", "B", "AB", "Residuals"), aov_typeIII[-1,1], aov_typeIII[-1,4])
colnames(fr_III)<-c("Source", "Sum Sq", "Pr(>F)")
full_fr<-cbind(fr_A, fr_B, fr_AB, fr_AB2, fr_III)
full_fr

## Table 18-21 - ANOVA and correlation tables from regression analysis with eight runs

con.A<-c(-1,1,-1,1,-1,1,1,1)
con.B<-c(-1,2,1,1,1,-1,-1)
con.AB<-con.A*con.B
d1<-data.frame(data_al, con.A, con.B, con.AB)
colnames(d1)<-c("resp", "A", "B", "AB")
fm_d1<-lm(resp~A*B, data=d1)
an_d1<-anova(fm_d1)

library(Hmisc)
corr_d1<-rcorr(as.matrix(d1))

d2<-data.frame(data_ac, con.A, con.B, con.AB)
colnames(d2)<-c("resp", "A", "B", "AB")
fm_d2<-lm(resp ~ A*B, data=d2)
an_d2<-anova(fm_d2)

corr_d2<-rcorr(as.matrix(d2))

## Table 24 - MANOVA table

y<-cbind(data_al, data_ac)
A1<-c(-1,1,-1,1,-1,1)
B1<-c(-1,-1,1,1,-1,1)
AB1<-A1*B1
man1<-manova(y~A1*B1)
summary(man1, test="Wilks")

## Plot 1 - MANOVA plot

fac<-c(-1,-2,1,1,-1,1)
fac<-as.factor(fac)
dat_man<-cbind(y,fac)
dat_man<-as.data.frame(dat_man)
attach(dat_man)
plot(data_al, data_ac, ylab="Tear strength across", xlab="Tear strength alongside", col=c("red","red","blue"))[fac]
legend(x="topright", title="Factor B", legend = levels(fac), col=c("red","red","blue"), pch=1)

# Table 27 - confidence intervals for effect estimates ANOVA eight runs (across)

A<-c(-1,1,-1,1,-1,1)
B<-c(-1,1,1,1,-1,1)
AB<-A*B
fm_an8<-lm(data_ac~A*B)
conf_an8<-cbind(fm_an8$coef, confint(fm_an8))
colnames(conf_an8)<-c("Estimate", "Lower Bound", "Upper Bound")
conf_an8
## Table 28 - confidence intervals for effect estimates ANOVA seven runs (across)

data_l7<-c(5479.67, 5559.57, 5406.87, 5493.90, 5460.20, 5472.43, 5665.47)
data_l7<-c(6030.8, 6354.67, 6329.10, 6013.77, 6251.47, 6276.53, 6194.00)

con.A7<-c(-1,-1,1,-1,1,-1,1)
con.B7<-c(-1,1,1,-1,1,1,1)
con.AB7<-con.A7*con.B7
d1_7<-data.frame(data_l7, con.A7, con.B7, con.AB7)
colnames(d1_7)<-c("resp", "A", "B", "AB")
fm_d1_7<-lm(resp~A*B, data=d1_7)

data_t7<-c(6030.8, 6354.67, 6329.10, 6013.77, 6251.47, 6276.53, 6194.00)
donames(d2_7)<-c("resp", "A", "B", "AB")
fm_d2_7<-lm(resp~A*B, data=d2_7)

conf_d2_7<-cbind(fm_d2_7$coef, confint(fm_d2_7))
colnames(conf_d2_7)<-c("Estimate", "Lower Bound", "Upper Bound")

## Table 29 - confidence intervals for coefficient estimates regression eight runs (across)

conf_d2<-cbind(fm_d2$coef, confint(fm_d2))
colnames(conf_d2)<-c("Estimate", "Lower Bound", "Upper Bound")

## Plot 2 - Residual analysis of candidate model

four_plot<-par(mfrow=c(2,2))
plot(fm_d2)
par(four_plot)

## Interaction plot (across) for candidate model

attach(d2)
interaction.plot(con.B, con.A, resp, main="interaction plot for regression with 8 runs",
                  xlab="factor A", trace.label=c("factor B"))

## Plot 4 - Contourplot (across)

x<-seq(-1,1,by=0.1)
y<-seq(-1,1,by=0.1)

model.fit1<-function(a,b) {
  6203.70 + 34.76 *a + 88.48 *b - 58.18 *a *b
}
z <- outer(x, y, model.fit1)
contour(x, y, z, ylab=c("Factor B"), xlab=c("Factor A"))
filled.contour(x, y, z, ylab="Factor B", xlab="Factor A", color.palette = colorRampPalette(c("white","orange","red")))